

## Recommendations for Primary Prevention of Skin Melanoma

Tõnis Tasa<sup>1</sup>, Mikk Puustusmaa<sup>1</sup>, Neeme Tõnisson<sup>1,2,4</sup>, Berit Kolk<sup>1</sup>, Peeter Padrik<sup>1,3,5</sup>

<sup>1</sup>Antegenes, Tartu, Estonia

<sup>2</sup>Institute of Genomics, University of Tartu, Tartu, Estonia

<sup>3</sup>Institute of Clinical Medicine, University of Tartu, Tartu, Estonia

<sup>4</sup>Center of Genetics, Tartu University Hospital, Tartu, Estonia

<sup>5</sup>Clinic of Hematology and Oncology, Tartu University Hospital, Tartu, Estonia

Corresponding Author:

Tõnis Tasa<sup>1</sup>

Email address: [tonis.tasa@antegenes.com](mailto:tonis.tasa@antegenes.com)

### Abstract

Melanoma (MEL) is an aggressive form of skin cancer, causing over 60,000 deaths every year and it is considered one of the fastest-growing cancer forms. Genome-wide association studies have identified numerous genetic variants (SNPs) independently associated with MEL. The effects of such SNPs can be combined into a single polygenic risk score (PRS). Stratification of individuals according to PRS could be introduced to the primary prevention of melanoma. Our aim was to combine PRS with health behavior recommendations to develop a personalized recommendation for primary prevention of melanoma.

Previously published PRS models for predicting the risk of melanoma were collected from the literature. Models were validated on the UK Biobank dataset consisting of a total of 487,410 quality-controlled genotypes with 3791 prevalent and 2345 incident cases. The best performing sex-specific models were selected based on the AUC in prevalent data and independently validated on an independent UKBB incident dataset for females and males separately. The best performing model included 28 SNPs. The C-index of the best performing model in the dataset was 0.59 (0.009) and hazard ratio (HR) per unit of PRS was 1.38 (standard error of log (HR) = 0.03) for both males and females.

We performed absolute risk simulations on the Estonian population and developed individual risk-based clinical follow-up recommendations. Both models were able to identify individuals with more than a 2-fold risk increase. The observed 10-year risks of developing melanoma for individuals in the 99th percentile exceeded the risk of individuals in the 1<sup>st</sup> percentile more than 4.5-fold.

We have developed a PRS-based recommendations pipeline for individual health behavior suggestions to support melanoma prevention.

## Introduction

Skin cancer is the most common form of cancer in Caucasians forming 40-50% of all cancer diagnoses in US (1, 2). Skin cancer can be divided into two types: non-melanoma skin cancers and melanoma. Latter being the most aggressive and lethal form of skin cancer (2). Over 280,000 new melanoma cases were diagnosed and more than 60,000 deaths were registered worldwide in 2018 (3). The number of melanoma incidences in the EU can reach up to 90,000 new cases per year, and it is considered one of the fastest-growing cancer forms in the EU (4). Melanoma is about 1.5 times more frequent in males than in females. The risk of melanoma rises with age and the average age at diagnosis is about 60 (5). Melanoma is also one of the most common types of cancer among adolescents and young adults (6). Melanoma is curable by surgical excision in the majority of cases if detected at an early stage (7).

The development of melanoma is complex, and it is interconnected with several environmental and genetic risk factors. Most cases of melanoma develop sporadically with no family history and approximately only 10% of individuals diagnosed with malignant melanoma have a first- or second-degree relative with a history of the disease (8, 9). The most significant environmental risk factor is exposure to ultraviolet (UV) radiation causing about 60-70% malignant melanomas (2, 10, 11). Solar UV radiation is 95% of UVA (315–400 nm) and it is also the primary source of radiation in tanning beds which can reach UVA doses up to 12-times that of the sun (2, 12). UVA leads to tumorigenesis primarily through oxidative stress-induced DNA damage as melanin in melanocytes acts as a photosensitizer causing diminished repair of UVA-induced oxidative damage (13, 14). However, the skin's response to UV radiation depends on many factors including the history of tanning, hair color, and skin type (2). In addition to phenotypic and environmental factors, genetics has a role in melanoma development. Several high penetrance susceptibility genes like CDKN2A, CDK4, BRCA1, BAP1, TERT, POT1, ACD, TERF2IP, and CXC genes and some moderate penetrance susceptibility genes like MC1R and MITF are known (15). Mutation in these genes confers to a high or moderate risk of developing melanoma. However, the frequency of mutations in these genes is rare, for example, in sporadic melanoma patients without personal or familial history of melanoma the probability of mutation in the CDKN2A gene is only about 1% (16). Therefore, combining the effects of many common genomic variants to predict melanoma is reasonable.

Multiple genome-wide association studies (GWAS) have identified dozens of loci and single-nucleotide polymorphisms (SNPs) associated with melanoma (17-26). These common genomic variants can be aggregated into a single polygenic risk score (PRS). In recent years, several different PRS models have been published and It has been shown that a PRS is strongly associated with melanoma risk allowing us to stratify patients by their cancer risk (27-31). Clinical usefulness of skin cancer PRS has been analyzed by M.R. Roberts and his colleagues and they concluded that PRS might already offer meaningful improvements to risk prediction (32).

The aim of this study is to evaluate the risk prediction performance of several published melanoma PRS and to assess its use as a risk stratification approach in the context of Estonia. We aim to use information from polygenic risk stratification to propose individual follow-up actions.

## Methods

### Participant data of UK Biobank

This study used genotypes from the UK Biobank cohort (obtained 07.11.2019) and made available to Antegenes under application reference number 53602. The data was collected, genotyped using either the UK BiLEVE or Affymetrix UK Biobank Axiom Array. Melanoma cases in the UK Biobank cohort were retrieved by the status of ICD-10 codes C43. We additionally included cases with self-reported UK Biobank code "1059".

Quality control steps and in detail methods applied in imputation data preparation have been described by the UKBB and made available at

[http://www.ukbiobank.ac.uk/wp-](http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf)

[content/uploads/2014/04/UKBiobank\\_genotyping\\_QC\\_documentation-web.pdf](http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf). We applied additional quality controls on autosomal chromosomes. First, we removed all variants with allele frequencies outside 0.1% and 99.9%, genotyping call rate <0.1, imputation (INFO) score <0.4 and Hardy-Weinberg equilibrium test p-value < 1E-6. Sample quality control filters were based on several pre-defined UK Biobank filters. We removed samples with excessive heterozygosity, individuals with sex chromosome aneuploidy, and excess relatives (> 10). Additionally, we only kept individuals for whom the submitted gender matched the inferred gender, and the genotyping missingness rate was below 5%.

Quality controlled samples were divided into prevalent and incident datasets for females and males separately. The prevalent datasets included MEL cases diagnosed before Biobank recruitment with 5 times as many controls without the diagnosis. Incident datasets included cases diagnosed in any of the linked databases after recruitment to the Biobank and all controls not included in the prevalent dataset. Prevalent datasets were used for identifying the best sex-specific candidate model and the incident datasets were used to obtain independent PRS effect estimates on MEL status.

### Model selection from candidate risk models

We searched the literature for PRS models in the public domain. The requirements for inclusion to the candidate set were the availability of the chromosomal location, reference and alternative allele, minor allele frequency, and an estimator for the effect size either as odds ratio (OR) or its logarithm (log-OR) specified for each genetic variant. In cases of iterative model developments on the same underlying

base data, we retained chronologically newer ones. The search was performed with Google Scholar and PubMed web search engines by working through a list of articles using the search ["Polygenic risk score" or "genetic risk score" and "melanoma"], and then manually checking the results for the inclusion criteria. We additionally pruned the PRS from multi-allelic, non-autosomal, non-retrievable variants based on bioinformatics re-analysis with Illumina GSA-24v1 genotyping.

PRSs were calculated as  $PRS = \sum_i^m \beta_i (\sum_{j=0}^2 \omega_{ij} x_j)$ , where  $\omega_{ij}$  is the probability of observing genotype  $j$ , where  $j \in \{0, 1, 2\}$  for the  $i$ -th SNP;  $m$  is the number of SNPs; and  $\beta_i$  is the effect size of the  $i$ -th SNP estimated in the PRS. The mean and standard deviation of PRS in the cohort were extracted to standardize individual risk scores to Gaussian. We tested the assumption of normality with the mean of 1000 Shapiro-Wilks test replications on a random subsample of 1000 standardized PRS values.

Next, we evaluated the relationship between MEL status and standardized PRS in the two sex-specific prevalent datasets with a logistic regression model to estimate the logistic regression-based odds ratio per 1 standard deviation of PRS ( $OR_{sd}$ ), its p-value, model Akaike information criterion (AIC) and Area Under the ROC Curve (AUC). The logistic regression model was compared to the null model using the likelihood ratio test and to estimate the Nagelkerke and McFadden pseudo- $R^2$ . We selected the candidate model with the highest AUC to independently assess risk stratification in the incident dataset.

## Independent performance evaluation of a polygenic risk score model

Firstly, we repeated the main analyses in the prevalent dataset. The main aim was to derive a primary risk stratification estimate, hazard ratio per 1 unit of standardized PRS ( $HR_{sd}$ ), using a right-censored and left-truncated Cox-regression survival model. The start of time interval was defined as the age of recruitment; follow-up time was fixed as the time of diagnosis. Scaled PRS was fixed as the only independent variable of MEL diagnosis status. 95% confidence intervals were created using the standard error of the log-hazard ratio.

Further, we evaluated the concordance between theoretical hazard ratio estimates derived with the continuous per unit PRS ( $HR_{sd}$ ) estimate and the hazard ratio estimates inferred empirically from data. For this, we binned the individuals by PRS to 5%-percentiles and estimated empiric hazard ratio of MEL directly between those classified in each bin and those within 40-60 PRS percentile. Theoretical estimates were derived from the relationship  $HR_{sd}^{\Phi^{-1}(a,b)}$ , where the exponent is the expected Gaussian value between two arbitrary percentiles  $a$  and  $b$  (bounded between 0 and 1,  $a < b$ ) of the Gaussian distribution,  $\Phi^{-1}(a,b) = \frac{(f(Q(b)) - f(Q(a)))}{(a - b)}$ , where  $Q(b)$  is the Gaussian quantile function on a percentile  $b$  and  $f(Q(b))$  is the Gaussian probability density function value at a quantile

function value. We compared the two approaches by using the Spearman correlation coefficient and the proportion of distribution-based  $HR_{sd}^{\Phi^{-1}(a,b)}$  estimates in empirical confidence intervals.

## Absolute risk estimation

Individual  $\tau$ -year (eg. 10-year) absolute risk calculations are based on the risk model developed by Pal Choudhury *et al* (33). Individual absolute risks are estimated for currently  $a$ -year old individuals in the presence of known risk factors ( $Z$ ) and their relative log hazard-ratio parameters ( $\beta$ ). 95% uncertainty intervals for the hazard ratio were derived using the standard error and z-statistic 95% quantiles  $CI_{HR} = \exp(\beta \pm 1.96 * se(HR))$ , where  $se(HR)$  is the standard error of the log-hazard ratio estimate. Risk factors have a multiplicative effect on the base-line hazard function. The model specifies the next  $\tau$ -year absolute risk for a currently  $a$ -year old individual as

$$\int_a^{a+\tau} \lambda_0(t) \exp(\beta^T Z) \exp\left(-\int_a^t [\lambda_0(u) \exp(\beta^T Z) + m(u)] du\right) dt,$$

where  $m(t)$  is age-specific mortality rate function and  $\lambda_0(t)$  is the baseline-hazard function,  $t \geq T$  and  $T$  is the time to onset of the disease. The baseline-hazard function is derived from marginal age-specific MEL incidence rates ( $\lambda_m(t)$ ) and distribution of risk factors  $Z$  in the general population ( $F(z)$ ).

$$\lambda_m(t) \approx \int \lambda_0(t) \exp(\beta^T Z) dF(z)$$

This absolute risk model allows disease background data from any country. In this analysis, we used Estonian background information. We calculated average cumulative risks using data from National Institute of Health Development of Estonia (34) that provides population average disease rates in age groups of 5-year interval. Sample sizes for each age group was acquired from Statistics Estonia for 2013-2016. Next, we assumed constant incidence rates for each year in the 5-year groups. Thus, incidence rates for each age group were calculated as  $IR = X_i / N_i$ , where  $X_i$  is the number of first-time cases at age  $t$  and  $N_i$  is the total number of women in this age group. Final per-year incidences were averaged over time range 2013-2016. Age- and sex-specific mortality data were retrieved from World Health Organization (35) and competing mortality rates were constructed by subtracting yearly age- and sex-specific disease mortality rates from general mortality rates. Melanoma mortality estimates were derived from Global Cancer Observatory (36).

We applied this model to estimate absolute risks for individuals in the 1<sup>st</sup>, 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup> and 99<sup>th</sup> PRS quantiles, eg. an individual on the 50<sup>th</sup> percentile would have a standardized PRS of 0. Confidence intervals for the absolute risk are estimated with the upper and lower confidence intervals of

the continuous per unit log-hazard ratio. Similarly, we used the absolute risk model to estimate lifetime risks (between ages 0 and 85) for the individuals in the same risk percentiles.

## Polygenic risk score based risk stratification

Next, we simulate cumulative PRS stratified lifetime risks using sex-specific MEL incidences in Estonian population by evaluating lifetime absolute risks for individuals in various PRS risk percentiles.

Our analysis first established the 10-year risk of a 50-year old female with a population average of PRS ("average individual") using the model by Pal Choudhury (33). Additionally, we assessed the differences in ages where individuals in various PRS risk percentiles attain 1 to 3-fold increases of risk compared to the 10-year risk of an average individual of the same sex and explore the dynamics of melanoma incidences between different sexes. Lastly, we added general health and observation recommendations with special caveats for groups specifically at high risk based on current practices of melanoma monitoring.

## Results

There was a total of 487,410 quality-controlled genotypes available for males and females in the complete UK Biobank cohort. Both genotype and phenotype data were available for a total of 458 696 individuals. This included 235 578 (234 408 controls, 1170 cases) cases for incident females and 13484 (11232 controls, 2252 cases) for prevalent females, and additional 9215 prevalent (7676 controls, 1539 cases) males and 200 419 (199 233 controls, 1186 cases) incident males.

Table 1. Comparison metrics of melanoma PRS models based on the prevalent UKBB dataset.

		ML23 (28)	ML29 (37)	ML33 (39)	ML36 (31)	ML37 (29)
Variants in original PRS		34	202	53	40	16
Variants included in our model		28	194	40	32	13
Males	AUC (SE)	0.580 (0.019)	0.573 (0.018)	0.552 (0.018)	0.565 (0.018)	0.566 (0.018)
	OR <sub>sd</sub> ((SE [log OR <sub>sd</sub> ]))	1.34 (0.03)	1.32 (0.03)	1.20 (0.03)	1.27 (0.03)	1.26 (0.03)
	AIC	8218.7	8227.1	8280.2	8247.1	8251.6
	McFadden / Nagelkerke Pseudo-R <sup>2</sup>	1.2% / 1.8%	1.1% / 1.7%	0.5% / 0.7%	0.9% / 1.3%	0.8% / 1.2%

Females	AUC (SE)	0.574 (0.015)	0.566 (0.014)	0.557 (0.015)	0.568 (0.015)	0.557 (0.015)
	OR <sub>sd</sub> ((SE [log OR <sub>sd</sub> ]))	1.31 (0.02)	1.27 (0.02)	1.22 (0.02)	1.27 (0.02)	1.23 (0.02)
	AIC	12039	12065	12099	12057	12091
	McFadden / Nagelkerke Pseudo-R <sup>2</sup>	1.1% / 1.6%	0.9% / 1.3%	0.6% / 0.9%	0.9% / 1.4%	0.6% / 1.0%

Altogether, we re-validated 5 PRS models (28, 30, 31, 37-39). We used only the PRS made up of previously known GWAS catalog SNPs and SNPs newly identified in their analysis as other compared PRS models included UKBB data in ML23 from Fritsche *et al.* (28). Normality assumption of the standardized PRS was not violated with any tested models (Shapiro-Wilks test p-values in ML23 = 0.43, ML29 = 0.50, ML33 = 0.35, ML36 = 0.13, ML37 = 0.40). The best performing model was selected based on AUC, OR<sub>sd</sub>, AIC and pseudo-R<sup>2</sup> metrics for both females and males. The ML23 model that was based on Fritsche *et al.* (28) performed the best (Table 1). The model's AUC under the ROC curve for the association between the PRS and MEL diagnosis was 0.580 (SE = 0.019) for males (Figure 1) and 0.574 (SE = 0.015) for females.

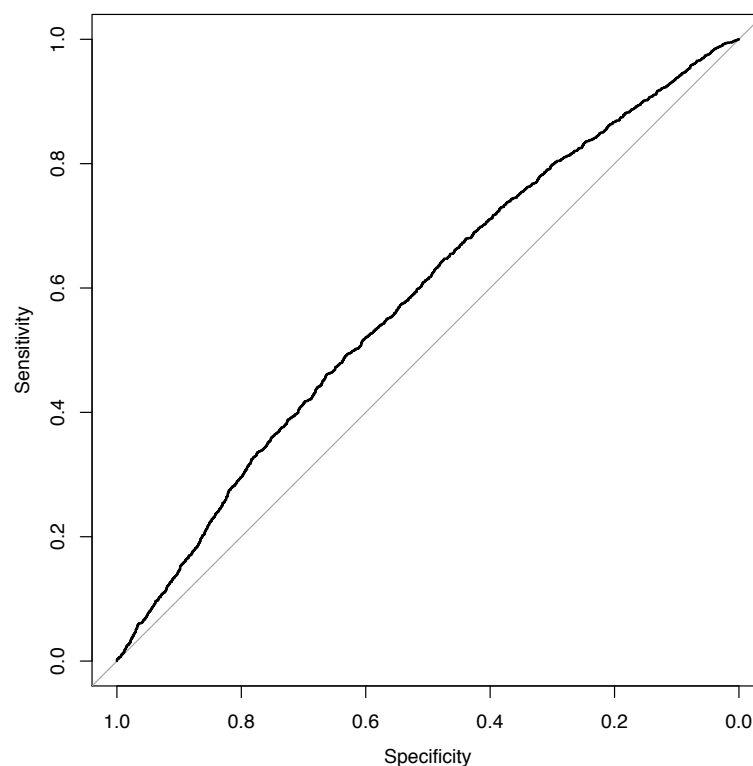
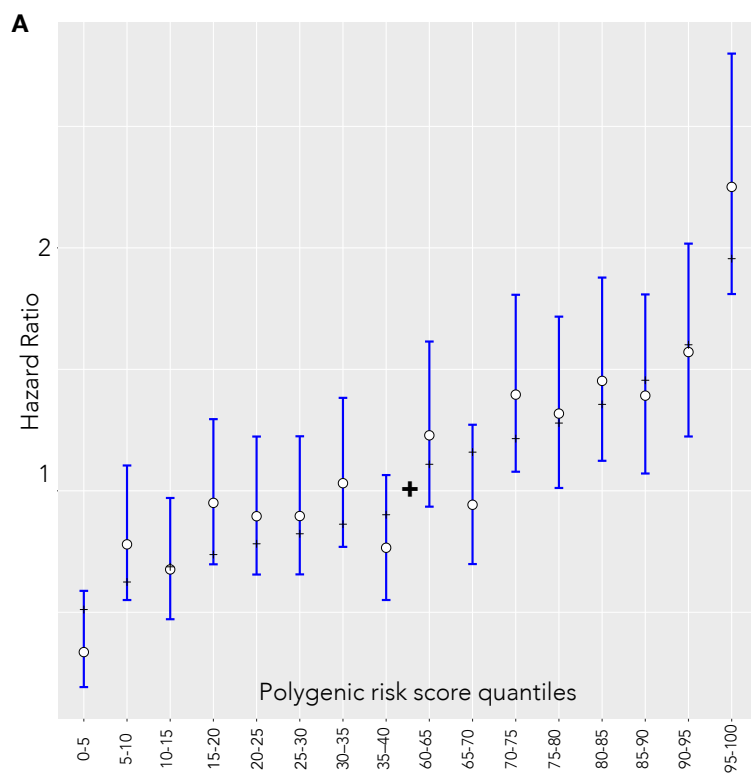


Figure 1. ROC plot of MEL cases and controls in UK Biobank prevalent dataset of males

Next, we evaluated the performance of the best performing ML23 model in an independent UKBB incident dataset with the main aim of estimating the hazard ratio per unit of PRS. Table 2 presents the performance estimation metrics in the incident dataset. Hazard ratio per 1 unit of standard deviation ( $HR_{sd}$ ) in model ML23 was 1.38 with standard error ( $\log(HR)$ ) = 0.03 for both males and females. The concordance index (C-index) of the survival model testing the relationship between PRS and MEL diagnosis status in the males' dataset was 0.587 (0.009) and again nearly completely equivalent in females.

Table 2. Performance metrics of the ML23 model in the incident UK Biobank dataset.

		$HR_{sd}$ (95% confidence interval)	C-index (SE)	-2 x log likelihood	Likelihood ratio test p-value
ML23 (5)	Males	1.38 (1.31-1.47)	0.587 (0.009)	120.3	< 2e-16
	Females	1.38 (1.3 – 1.46)	0.588 (0.009)	117.1	< 2e-16





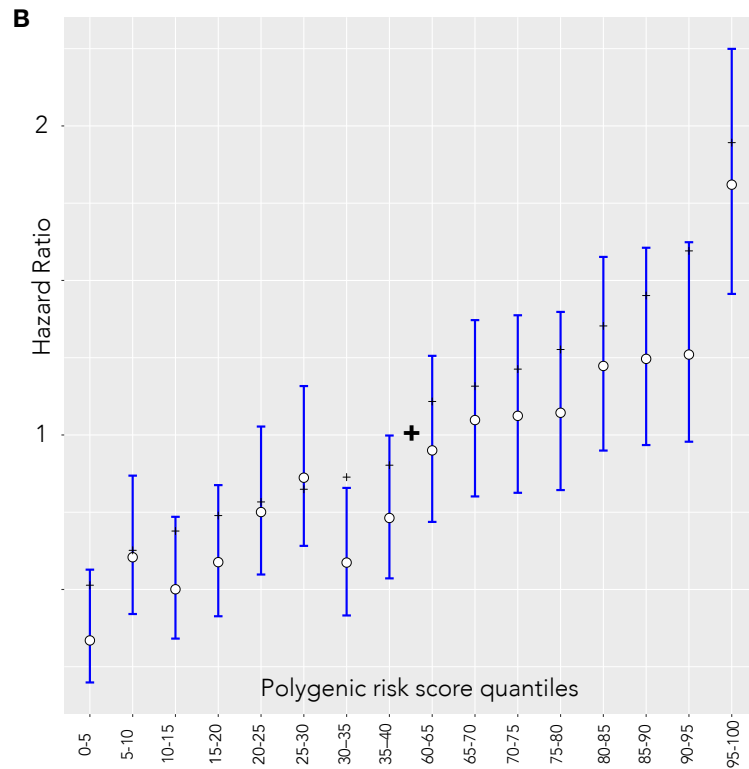


Figure 2. Hazard ratio estimates between quantiles 40-60 of the PRS and categorized 5% bins in the UKBB incident dataset. White dots and blue lines represent empirically estimated hazard ratio estimates and corresponding confidence intervals. Black dashes represent the theoretical hazard ratio for the 5%-quantile bins derived from the hazard ratio of per unit PRS. (A) Males (B) Females

Hazard ratio estimates compared to individuals in 40-60 percentile of PRS are visualized in Figure 2. In females' dataset (panel B), the theoretical hazard ratio matched in empirical estimate's confidence intervals in 15 out of 16 comparisons (Spearman correlation 0.973). In males' dataset (panel A), the intervals matched in all comparisons (Spearman correlation 0.948).

## Risk stratification with a melanoma polygenic risk score

We used a model by Pal Choudhury et al. (33) to derive individual 10-year risks and specified  $F(z)$  as the distribution of PRS estimates in the whole UKBB cohort. The log-hazard ratio ( $\beta$ ) is based on the sex-specific estimate of the log-hazard ratio in the ML23 model of the incident UKBB dataset. Age-specific MEL incidence and competing mortality rates provided the background for MEL incidences in the Estonian population.

In the Estonian population, the absolute risk of developing melanoma in the next 10 years among 50-year old men in the 1<sup>st</sup> percentile is 0.08% (0.07% - 0.10%) and 0.37% (0.33% - 0.42%) in the 99<sup>th</sup>

percentile (0.12% (0.10% - 0.14%) and 0.53% (0.47% - 0.60%) for women, respectively). At age 70, corresponding risks for the same percentiles of men become 0.20% (0.18% - 0.24%) and 0.93% (0.83% - 1.04%). As opposed younger ages, at age 70 the 10-year risk of females is lower than of males: 0.17% (0.14% - 0.20%) and 0.75% (0.67% - 0.84%) respectively in the most extreme percentiles. The relative risks between the most extreme percentiles are therefore around 4.4-fold. Competing risks accounted cumulative risks of females in the 99th percentile surpass 2.13% (1.89% - 2.40%) by age 85 but remain at 0.47% (0.41% - 0.55%) in the 1<sup>st</sup> percentile (Figure 3). Equivalent values for women in Estonia exceed those of men: 2.74% (2.44% – 3.07%) and 0.62% (0.53% - 0.72%) in the most extreme percentiles.

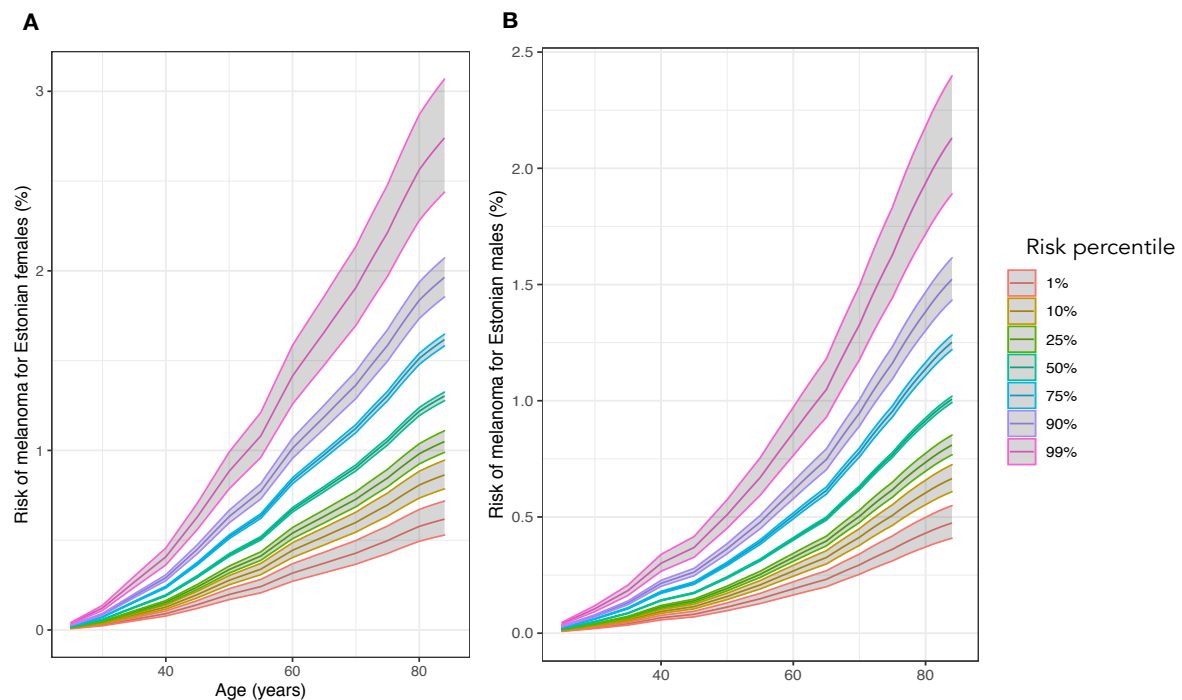


Figure 3. Cumulative risks (%) of melanoma between ages 20 to 85 in various risk percentiles. (A) Females (B) Males

A genetically average 40-year old has a 10-year risk of melanoma equaling 0.1% in males (Figure 4 panel A) and 0.218% in females (Figure 4 panel B). A 30-year-old male in the 99<sup>th</sup> percentile (26 for females) has a larger risk than an average 40-year-old. At the same time, males in the 1<sup>st</sup> percentile attain this risk by age 54. Females in the lowest 5% do not attain the average 40-year old's absolute average risk even by age 70. More than 1% of males and females have a 2-fold risk compared to average at age 40. The 10-year risk of a 40-year-old male doubles by age 56 in genetically average males and triples by age 63. Importantly, such risk increases do not occur in females that have a larger proportion of incidences in younger ages. The 1<sup>st</sup> percentile of males only triples the population average 40-year-old's risk by age 70. This does not occur in females.

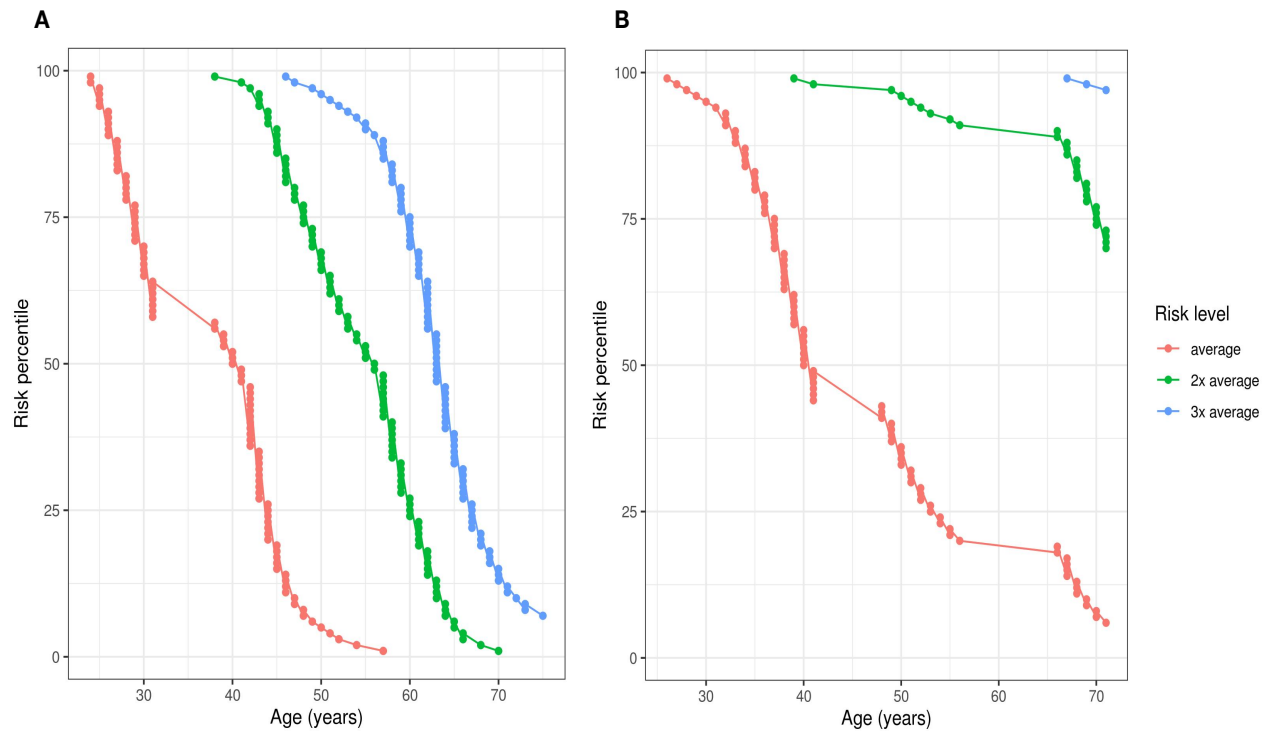


Figure 4. Ages when Estonian individuals in different risk percentiles attain 1-3 fold multiples of 10-year risk compared to 40-year olds with population average PRS (Risk level: "average"). (A) Males (B) Females,

No population screening approaches have been instituted to track and screen melanoma. Early detection can be supported with regular self-examinations and health behavior that apply to all. Also, we recommend once-a-year dermatology visits with individuals having greater than 2-fold risk increases compared to those with average PRS.

1. Recommendations for all
  - a. Watch for abnormal moles and perform a skin self-exam every month for any new changes. If you spot anything that just does not look right, get it checked by a dermatologist as soon as possible.
  - b. Limit your exposure to ultraviolet (UV) radiation: protect your skin by using a broad-spectrum sunscreen with an SPF 30 or higher
  - c. Avoid using tanning beds and sunlamps.
2. Relative risk increases over 2-fold
  - a. See a dermatologist once a year for a full-body skin exam.

## Discussion

A record number of new skin cancers cases are seen worldwide causing a significant burden to the health care system (2, 40). Melanoma is the most lethal form of skin cancer and in the U.S., the cost of melanoma treatment is about \$3.3 billion each year (2, 41). The risk of melanoma rises with age and the average age at diagnosis is about 60. With an aging population, the number of melanoma cases will likely remain high for the near future. It has been shown that genetics has a role in melanoma development. Common genomic variants can help identify people at high risk for melanoma, particularly those who lack traditional risk factors (31).

In this study, we validated different publicly available PRS models to find a sex-specific model that could be used to design a novel risk-based screening strategy for melanoma. Our best-performing model, named ML23, was a pruned version of Fritsche et al (28) PRS model containing a total of 28 SNPs. In the UK Biobank incident dataset, the AUC was 0.58 and the estimated hazard ratio (HR) for overall MEL per unit PRS was 1.38 for both sexes. Our model was able to identify individuals with more than 4-fold PRS-based risk differences between the extremes.

We also developed personalized health behavior recommendations based on PRS result and overall guidelines for melanoma prevention. First, as nevi are one of the strongest risk factors for melanoma (42) we recommend to watch for abnormal moles and perform a skin self-exam every month for any new changes. Second, it is estimated that 60–70% of melanomas are caused by ultraviolet (UV) radiation (2, 11). Thus, primary prevention strategies that reduce sun exposure and encourage increased sun protection are important for reducing melanoma incidence and mortality (10, 43-45). Therefore, we recommend to limit exposure to ultraviolet to all patients and suggest protecting their skin by using a broad-spectrum sunscreen with an SPF 30 or higher. Besides, as tanning beds can reach UVA doses 12-times that of the sun (12) we recommend all patents to avoid using tanning beds and sunlamps if possible, regardless of the fact that skin's response to UV radiation depends on many factors including history of tanning, hair color, skin type and genetic background (2). For individuals, whose relative risk is increased over 2-folds, we recommend full-body skin exam by a dermatologist once a year. Survival from melanoma is strongly related to tumor thickness and there is strongest evidence that whole-body skin examination reduces the incidence of thick melanoma (43). Our approach is easily adaptable to other nationalities by using population background information data of other genetically similar populations.

In conclusion, our PRS based model identifies individuals at more than 2-fold risk and provides recommendations for melanoma prevention.

## Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Reference Number 53602 and with support by EIT Health The Digital Sandbox program.

## Ethical approval

UK Biobank: The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee (UK Biobank reference: 16/NW/0274). All participants provided written informed consent to participate in the UK Biobank study.

## Statement of data availability

Individual level genotype and phenotype data from UK Biobank can not be explicitly shared. The UK Biobank Resource was used under Application Reference Number 53602. New users can request access to UK Biobank from <http://www.ukbiobank.ac.uk/resources/>.

## Funding

OÜ Antegenes has received a grant from the EIT Health The Digital Sandbox program and additional Innovation Voucher funding meant for business development of small and medium sized Estonian enterprises.

## References

1. Alam M, Nanda S, Mittal BB, Kim NA, Yoo S. The use of brachytherapy in the treatment of nonmelanoma skin cancer: a review. *Journal of the American Academy of Dermatology*. 2011;65(2):377-88.
2. Sample A, He YY. Mechanisms and prevention of UV-induced melanoma. *Photodermatol Photoimmunol Photomed*. 2018;34(1):13-24.
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018;68(6):394-424.
4. Commission E. Melanoma skin cancer. 2017.
5. Apalla Z, Lallas A, Sotiriou E, Lazaridou E, Ioannides D. Epidemiological trends in skin cancer. *Dermatology practical & conceptual*. 2017;7(2):1-6.
6. Iannacone MR, Youlden DR, Baade PD, Aitken JF, Green AC. Melanoma incidence trends and survival in adolescents and young adults in Queensland, Australia. *International journal of cancer*. 2015;136(3):603-9.
7. Orzan OA, Sandru A, Jecan CR. Controversies in the diagnosis and treatment of early cutaneous melanoma. *J Med Life*. 2015;8(2):132-41.
8. Hayward NK. Genetics of melanoma predisposition. *Oncogene*. 2003;22(20):3053-62.
9. Debniak T. Familial malignant melanoma - overview. *Hereditary cancer in clinical practice*. 2004;2(3):123-9.
10. Armstrong BK, Kricker A. How much melanoma is caused by sun exposure? *Melanoma Res*. 1993;3(6):395-401.

11. Koh HK, Geller AC, Miller DR, Grossbart TA, Lew RA. Prevention and early detection strategies for melanoma and skin cancer. Current status. *Archives of dermatology*. 1996;132(4):436-43.
12. van Weelden H, de Gruijl FR, van der Putte SC, Toonstra J, van der Leun JC. The carcinogenic risks of modern tanning equipment: is UV-A safer than UV-B? *Arch Dermatol Res*. 1988;280(5):300-7.
13. de Gruijl FR. Photocarcinogenesis: UVA vs. UVB radiation. *Skin Pharmacol Appl Skin Physiol*. 2002;15(5):316-20.
14. Wang HT, Choi B, Tang MS. Melanocytes are deficient in repair of oxidative DNA damage and UV-induced photoproducts. *Proc Natl Acad Sci U S A*. 2010;107(27):12180-5.
15. Potrony M, Badenas C, Aguilera P, Puig-Butille JA, Carrera C, Malvehy J, et al. Update in genetic susceptibility in melanoma. *Ann Transl Med*. 2015;3(15):210.
16. Harland M, Cust AE, Badenas C, Chang YM, Holland EA, Aguilera P, et al. Prevalence and predictors of germline CDKN2A mutations for melanoma cases from Australia, Spain and the United Kingdom. *Hereditary cancer in clinical practice*. 2014;12(1):20.
17. Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nature genetics*. 2008;40(7):838-40.
18. Bishop DT, Demenais F, Iles MM, Harland M, Taylor JC, Corda E, et al. Genome-wide association study identifies three loci associated with melanoma risk. *Nature genetics*. 2009;41(8):920-5.
19. Amos CI, Wang LE, Lee JE, Gershenwald JE, Chen WV, Fang S, et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Human molecular genetics*. 2011;20(24):5012-23.
20. Barrett JH, Iles MM, Harland M, Taylor JC, Aitken JF, Andresen PA, et al. Genome-wide association study identifies three new melanoma susceptibility loci. *Nature genetics*. 2011;43(11):1108-13.
21. Macgregor S, Montgomery GW, Liu JZ, Zhao ZZ, Henders AK, Stark M, et al. Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3. *Nature genetics*. 2011;43(11):1114-8.
22. Iles MM, Law MH, Stacey SN, Han J, Fang S, Pfeiffer R, et al. A variant in FTO shows association with melanoma risk not due to BMI. *Nature genetics*. 2013;45(4):428-32, 32e1.
23. Gudbjartsson DF, Sulem P, Stacey SN, Goldstein AM, Rafnar T, Sigurgeirsson B, et al. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nature genetics*. 2008;40(7):886-91.
24. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, et al. Two newly identified genetic determinants of pigmentation in Europeans. *Nature genetics*. 2008;40(7):835-7.
25. Duffy DL, Zhu G, Li X, Sanna M, Iles MM, Jacobs LC, et al. Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. *Nature communications*. 2018;9(1):4774.
26. Law MH, Bishop DT, Lee JE, Brossard M, Martin NG, Moses EK, et al. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. *Nature genetics*. 2015;47(9):987-95.
27. Gu F, Chen TH, Pfeiffer RM, Fagnoli MC, Calista D, Ghiorzo P, et al. Combining common genetic variants and non-genetic risk factors to predict risk of cutaneous melanoma. *Human molecular genetics*. 2018;27(23):4145-56.
28. Fritsche LG, Beesley LJ, VandeHaar P, Peng RB, Salvatore M, Zawistowski M, et al. Exploring various polygenic risk scores for skin cancer in the phenomes of the Michigan genomics initiative and the UK Biobank with a visual catalog: PRSWeb. *PLoS genetics*. 2019;15(6):e1008202.
29. Fritsche LG, Gruber SB, Wu Z, Schmidt EM, Zawistowski M, Moser SE, et al. Association of polygenic risk scores for multiple cancers in a phenome-wide study: results from the Michigan Genomics Initiative. *The American Journal of Human Genetics*. 2018;102(6):1048-61.
30. Cho HG, Ransohoff KJ, Yang L, Hedlin H, Assimes T, Han J, et al. Melanoma risk prediction using a multilocus genetic risk score in the Women's Health Initiative cohort. *Journal of the American Academy of Dermatology*. 2018;79(1):36-41.e10.
31. Cust AE, Drummond M, Kanetsky PA, Australian Melanoma Family Study I, Leeds Case-Control Study I, Goldstein AM, et al. Assessing the Incremental Contribution of Common Genomic Variants to Melanoma Risk Prediction in Two Population-Based Studies. *The Journal of investigative dermatology*. 2018;138(12):2617-24.

32. Roberts MR, Asgari MM, Toland AE. Genome-wide association studies and polygenic risk scores for skin cancer: clinically useful yet? *The British journal of dermatology*. 2019.
33. Pal Choudhury P, Maas P, Wilcox A, Wheeler W, Brook M, Check D, et al. iCARE: An R package to build, validate and apply absolute risk models. *PloS one*. 2020;15(2):e0228198.
34. HSaHR [Internet]. 2020.
35. Organization WH. Global Health Observatory indicator views 2018 [Available from: [https://apps.who.int/gho/data/node.imr.LIFE\\_0000000029?lang=en](https://apps.who.int/gho/data/node.imr.LIFE_0000000029?lang=en)].
36. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International journal of cancer*. 2019;144(8):1941-53.
37. Gu F, Chen T-H, Pfeiffer RM, Fargnoli MC, Calista D, Ghiorzo P, et al. Combining common genetic variants and non-genetic risk factors to predict risk of cutaneous melanoma. *Human molecular genetics*. 2018;27(23):4145-56.
38. Fang S, Han J, Zhang M, Wang LE, Wei Q, Amos CI, et al. Joint effect of multiple common SNPs predicts melanoma susceptibility. *PloS one*. 2013;8(12):e85642.
39. Kypreou KP, Stefanaki I, Antonopoulou K, Karagianni F, Ntritsos G, Zaras A, et al. Prediction of melanoma risk in a Southern European population based on a weighted genetic risk score. *Journal of Investigative Dermatology*. 2016;136(3):690-5.
40. Riker AI, Zea N, Trinh T. The epidemiology, prevention, and detection of melanoma. *Ochsner J*. 2010;10(2):56-65.
41. Guy GP, Jr., Machlin SR, Ekwueme DU, Yabroff KR. Prevalence and costs of skin cancer treatment in the U.S., 2002-2006 and 2007-2011. *Am J Prev Med*. 2015;48(2):183-7.
42. Olsen CM, Carroll HJ, Whiteman DC. Estimating the attributable fraction for cancer: A meta-analysis of nevi and melanoma. *Cancer prevention research (Philadelphia, Pa)*. 2010;3(2):233-45.
43. Aitken JF, Elwood M, Baade PD, Youl P, English D. Clinical whole-body skin examination reduces the incidence of thick melanomas. *International journal of cancer*. 2010;126(2):450-8.
44. Breitbart EW, Waldmann A, Nolte S, Capellaro M, Greinert R, Volkmer B, et al. Systematic skin cancer screening in Northern Germany. *Journal of the American Academy of Dermatology*. 2012;66(2):201-11.
45. Glanz K, Volpicelli K, Jepson C, Ming ME, Schuchter LM, Armstrong K. Effects of tailored risk communications for skin cancer prevention and detection: the PennSCAPE randomized trial. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2015;24(2):415-21.